

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

A0A

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 05-324473

(43)Date of publication of application : 07.12.1993

(51)Int.Cl.

G06F 12/08

G06F 12/08

(21)Application number : 04-157467

(71)Applicant : NEC CORP

(22)Date of filing : 25.05.1992

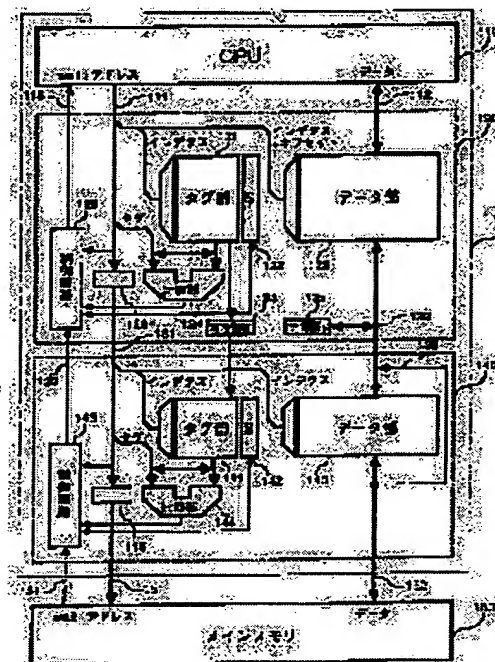
(72)Inventor : NAKAYAMA TAKASHI

(54) CACHE MEMORY SYSTEM AND MICROPROCESSOR UNIT

(57)Abstract:

PURPOSE: To and improves the performance at the time of making a CPU and the cache memory into one LSI.

CONSTITUTION: This system is composed of a CPU 110, direct map system first order cache memory 120, direct map system second-order cache memory 140 and main memory 160. The cache memories 120 and 140 are integrated in the same LSI as the CPU 110 and operated at high speed. The contents of the first-order and second-order cache memories 120 and 140 are controlled so as not to be overlapped. The index of the second-order cache memory is hashed by a hierarchical exclusive OR circuit 146. Thus, the small-area and high-performance microprocessor can be easily constructed by adding the small-scale and high-speed second-order cache memory 140 rather than by doubling the capacity of the first-order cache memory 120.



LEGAL STATUS

[Date of request for examination] 25.04.1996

[Date of sending the examiner's decision of rejection] 21.07.1998

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3262182

[Date of registration] 21.12.2001

[Number of appeal against examiner's decision of rejection] 10-13165

[Date of requesting appeal against examiner's decision of rejection] 20.08.1998

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平5-324473

(43)公開日 平成5年(1993)12月7日

(51)Int.Cl.⁵

G 0 6 F 12/08

識別記号

3 1 0 Z 7232-5B

庁内整理番号

F 7232-5B

F I

技術表示箇所

審査請求 未請求 請求項の数5(全 13 頁)

(21)出願番号

特願平4-157467

(22)出願日

平成4年(1992)5月25日

(71)出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72)発明者 中山 貴司

東京都港区芝五丁目7番1号 日本電気株式会社内

(74)代理人 弁理士 菅野 中

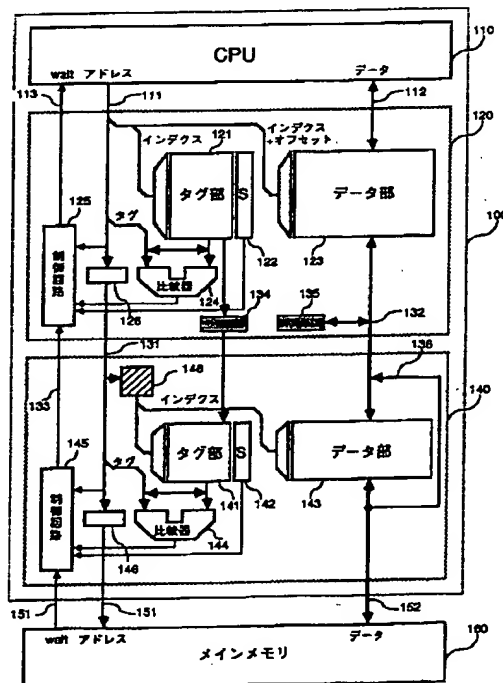
(54)【発明の名称】 キャッシュメモリ方式及びマイクロプロセッサ装置

(57)【要約】

【目的】 CPUとキャッシュメモリをひとつのLSI化するときに、占有面積が小さく、かつ性能の良いキャッシュメモリを提供する。

【構成】 CPU110とダイレクトマップ方式の1次キャッシュメモリ120とダイレクトマップ方式の2次キャッシュメモリ140とメインメモリ160から構成される。キャッシュメモリ120、140は、CPU110と同一LSIに集積され、高速に動作する。1次キャッシュメモリ120と2次キャッシュメモリ140の内容に重複がないように制御する。2次キャッシュメモリのインデックスは階層的な排他的論理和回路146によってハッシュされている。

【効果】 1次キャッシュメモリの容量を倍にするよりも、小規模かつ高速な2次キャッシュメモリを付加した方が、小面積で高性能なマイクロプロセッサが構築できる。



1

【特許請求の範囲】

【請求項1】 第1のキャッシュメモリと第2のキャッシュメモリとを有するキャッシュメモリ方式であって、各キャッシュメモリは、データ用メモリと、タグ用メモリと、比較器と、ヒット発生回路とを含み、データ用メモリは、ブロック単位にデータを記憶するものであり、タグ用メモリは、前記データ用メモリの各ブロックに記憶したデータがアドレス空間のどこに位置しているものかを示す情報（タグ）を記憶したものであり、比較器は、アドレス中のアドレスタグと前記タグ用メモリの出力を照合するものであり、ヒット発生回路は、前記比較器の内容に基づいてヒット信号を発生するものであり、内部からの読み出しアクセスに対して前記第1のキャッシュメモリ手段をアクセスし、前記第1のキャッシュメモリ手段でヒット信号が発生しなければ前記第2のキャッシュメモリ手段をアクセスし、前記第2のキャッシュメモリ手段でヒット信号が発生すれば前記第1のキャッシュメモリ手段の該当ブロックと前記第2のキャッシュメモリ手段の該当ブロックの内容を交換し、前記第2のキャッシュメモリ手段でヒット信号が発生しなければ外部メモリをアクセスすることを特徴とするキャッシュメモリ方式。

【請求項2】 前記第1のキャッシュメモリ手段のミスによるアクセスにおいて、アドレスの一部の排他的論理和をとって前記第2のキャッシュメモリ手段のタグ用メモリとデータ用メモリのアドレスとしたことを特徴とする請求項1に記載のキャッシュメモリ方式。

【請求項3】 前記第2のキャッシュメモリ手段のメモリ容量が前記第1のキャッシュメモリ手段のメモリ容量以下としたことを特徴とする請求項2に記載のキャッシュメモリ方式。

【請求項4】 請求項2に記載のキャッシュメモリ方式をもつ第1及び第2のキャッシュメモリ手段と中央処理装置とを同一のLSIチップに集積し、前記第1のキャッシュメモリ手段と前記第2のキャッシュメモリ手段の間の転送をクロックサイクル毎に行うことを特徴とするマイクロプロセッサ装置。

【請求項5】 請求項3に記載のキャッシュメモリ方式をもつ第1及び第2のキャッシュメモリ手段と中央処理装置とを同一のLSIチップに集積し、前記第1のキャッシュメモリ手段と前記第2のキャッシュメモリ手段の間の転送をクロックサイクル毎に行うことを特徴とするマイクロプロセッサ装置。

【発明の詳細な説明】

【0001】

2

【産業上の利用分野】 本発明は、キャッシュメモリ方式、特に、複数のダイレクトマップ方式のキャッシュメモリを階層的に接続したキャッシュメモリ方式及びマイクロプロセッサ装置に関する。

【0002】

【従来の技術】 近年の半導体装置の進歩をみると、マイクロプロセッサのクロック周波数が年々向上するのに対し、メインメモリとなるDRAMやROMのアクセス時間はそれほど短くならない。このスピードギャップを埋めるために、高速・小容量のメモリで構成するキャッシュメモリ装置をプロセッサとメインメモリの間に設ける方策が、よく採用されている。

【0003】 キャッシュメモリの基本方式は後述の参考1で詳しく説明されている。メインメモリをあらかじめ一定の容量（普通16Byte程度）にブロック分けし、これらの一部のブロックをキャッシュメモリに格納する。キャッシュメモリ内のブロックをエントリと呼ぶ。

【0004】 キャッシュメモリの各エントリは、3つの部分から構成される。（1）データを格納するデータ部、（2）どのアドレスにあるブロックが書込まれているかを示す情報（タグと呼ぶ）を格納するタグ部、（3）そのエントリに有効なデータが格納されているか否かを示すステータスフラグ、である。方式によってはステータスに別の意味を持たせることもある。

【0005】 キャッシュメモリの構成には、ダイレクトマップ方式、セット・アソシアティブ方式、フル・アソシアティブ方式の3通りがある。

【0006】 ①ダイレクトマップ方式

RAMで構成するタグ部、ステータスフラグ、データ部を1組ずつ持つ。アドレスの下位ビット（インデクス）をアドレスにしてRAM（タグ部とデータ部）をアクセスする。タグ部からの出力とアドレスの上位ビット（タグ）が等しく、ステータスフラグの内容が「有効」であれば、そのエントリのデータ部は有効（ヒット）である。

【0007】 ②セット・アソシアティブ方式

ダイレクトマップ方式のRAMをN組（普通2または4組）持ち、並列にアクセスする。ヒットした組がひとつでもあれば、ヒットした組のデータ部の出力が選択される。

【0008】 ③フル・アソシアティブ方式

各エントリのタグ部に比較器を持ち、アドレスとタグ部の内容を直接比較する。普通、CAM（Content Addressable Ram）で構成する。

【0009】 上記3方式のヒット率（ヒット回数／アクセス回数）は、①<②<③の順で高い。しかし、キャッシュメモリのアクセス時間も、①<②<③の順で長くなる。LSI化したときの面積は、①<②<③の順で大きくなる。特に、方式③は、面積の増加が大きく、ヒット

10

20

30

40

50

3

しなかったとき（ミスしたとき）の処理が複雑など、欠点が多い。

【0010】最近のRISC (Reduced Instruction Set Computer) CPUでは、キャッシュメモリのアクセス時間がクロック周波数に直接影響するため、①（ダイレクトマップ方式）を採用するものが多い。

【0011】①②方式のミス率（1-ヒット率）の違いを図8（参考文献1の図8、12のデータに基づく）に示す。上記方式①または②では、異なるアドレスが同一のエントリを使うよう強制されるためのミス（コンフリクト・ミスと呼ぶ）が発生するため、方式③よりもミス率が増加する。

【0012】もうひとつ図8から読取れるのは、キャッシュメモリの容量が大きくなるにしたがってコスト/パフォーマンスが低下していることである。キャッシュ容量が2倍になるにしたがって、ミス率は一定の割合で減少（0.7~0.8倍）している。一方、キャッシュメモリをLSI化したときの面積は、容量にほぼ比例している。

【0013】そのため、キャッシュメモリとCPUを1チップLSI化するには、省面積でかつミス率の低いキャッシュメモリ方式が望まれている。

【0014】もうひとつの従来例を挙げる。ダイレクトマップ方式（方式①：高速だがミス率が高い）のキャッシュメモリに、小規模なフルアソシアティブ方式（方式③：低速だがミスが低い）のキャッシュメモリを付加した、効率の良いキャッシュメモリ方式が後述の参考文献2で提案されている。

【0015】ダイレクトマップ方式キャッシュ（1次キャッシュ）がミスした場合、2次キャッシュとしてフルアソシアティブ方式キャッシュ（彼等はヴィクティムキャッシュと呼んでいる）をアクセスする。

【0016】2次キャッシュがミスした場合はメインメモリをアクセスする。どちらのキャッシュメモリもCPUと同一のLSIチップに組込まれているので、1次キャッシュと2次キャッシュの間の転送は高速（1クロックサイクル）で行なわれる。

【0017】ダイレクトマップ方式キャッシュメモリのコンフリクト・ミスの大半が2次キャッシュにヒットするので、1次キャッシュを2倍の容量にするのと同程度のミス率低減効果が、4~8エントリの2次キャッシュで得られる。

【0018】参考文献2では1次キャッシュと2次キャッシュの内容に重複が無いように制御することで、少ないエントリ数の2次キャッシュで効果を上げている。

【0019】この方式の欠点は、2次キャッシュにフルアソシアティブ方式を用いていることである。フルアソシアティブ方式は、各エントリに比較器を持つため面積の増加が著しい。また、ミスした場合にどのエントリを

4

入替えるかを定める制御論理（通常はLRU: Least Recent Used）が複雑であり、テストも難しい。ただし、速度が遅いという欠点だけは、1次キャッシュがミスしたときだけ使用されるため、あまり問題ではない。

【0020】参考文献1:「Computer Architecture: A Quantitative Approach」John L. Hennessy & David A. Patterson, 1990, Morgan Kaufmann Publishers Inc.

【0021】参考文献2:「Improving Direct-Mapped Cache Performance by the Addition of a Small Fully-Associative Cache and Prefetch Buffer」Norman P. Jouppi, 1990, IEEE International Symposium on Computer Architecture.

【0022】

【発明が解決しようとする課題】上述した従来の3種類のキャッシュメモリ方式の欠点をまとめる。

①ダイレクトマップ方式では、高速であるが、ミス率が高い。面積も小さい。

②セットアソシアティブ方式では、速度、面積では①に劣り、ミス率では③に劣る。

③フルアソシアティブ方式では、ミス率は一番低い。しかし、速度は遅く、面積が著しく大きい。制御論理も複雑であり、テストも難しい。

【0023】もうひとつのヴィクティムキャッシュ方式は、ダイレクトマップ方式①の高速性を活かしたまま、フルアソシアティブ方式③の低ミス率を実現する優れた方式である。ただ、方式③の面積、複雑といった欠点は有しているため、改良の余地がある。

【0024】本発明の目的は、CPUとキャッシュメモリを一体にLSI化するとき、占有面積が小さく、かつ性能の良いキャッシュメモリ方式を提供するとともに、このキャッシュメモリ方式を使用したマイクロプロセッサ装置を提供することにある。

【0025】

【課題を解決するための手段】前記目的を達成するため、本発明に係るキャッシュメモリ方式は、第1のキャッシュメモリと第2のキャッシュメモリとを有するキャッシュメモリ方式であって、各キャッシュメモリは、データ用メモリと、タグ用メモリと、比較器と、ヒット発生回路とを含み、データ用メモリは、ブロック単位にデータを記憶するものであり、タグ用メモリは、前記データ用メモリの各ブロックに記憶したデータがアドレス空間のどこに位置しているものかを示す情報（タグ）を記

憶したものであり、比較器は、アドレス中のアドレスタグと前記タグ用メモリの出力を照合するものであり、ヒット発生回路は、前記比較器の内容に基づいてヒット信号を発生するものであり、内部からの読み出しアクセスに対して前記第1のキャッシュメモリ手段をアクセスし、前記第1のキャッシュメモリ手段でヒット信号が発生しなければ前記第2のキャッシュメモリ手段をアクセスし、前記第2のキャッシュメモリ手段でヒット信号が発生すれば前記第1のキャッシュメモリ手段の該当ブロックと前記第2のキャッシュメモリ手段の該当ブロックの内容を交換し、前記第2のキャッシュメモリ手段でヒット信号が発生しなければ外部メモリをアクセスするものである。

【0026】また、前記第1のキャッシュメモリ手段のミスによるアクセスにおいて、アドレスの一部の排他的論理和をとって前記第2のキャッシュメモリ手段のタグ用メモリとデータ用メモリのアドレスとしたものである。

【0027】また、前記第2のキャッシュメモリ手段のメモリ容量が前記第1のキャッシュメモリ手段のメモリ容量以下としたものである。

【0028】また、本発明に係るマイクロプロセッサ装置は、第1のキャッシュメモリ手段のミスによるアクセスにおいて、アドレスの一部の排他的論理和をとって第2のキャッシュメモリ手段のタグ用メモリとデータ用メモリのアドレスとしたキャッシュメモリ方式をもつ第1及び第2のキャッシュメモリ手段と中央処理装置とを同一のLSIチップに集積し、前記第1のキャッシュメモリ手段と前記第2のキャッシュメモリ手段の間の転送をクロックサイクル毎に行うものである。

【0029】また、本発明に係るマイクロプロセッサ装置は、第2のキャッシュメモリ手段が第1のキャッシュメモリ手段のメモリ容量以下としたキャッシュメモリ方式をもつ第1及び第2のキャッシュメモリ手段と中央処理装置とを同一のLSIチップに集積し、前記第1のキャッシュメモリ手段と前記第2のキャッシュメモリ手段の間の転送をクロックサイクル毎に行うものである。

【0030】

【作用】上述した従来のヴィクティム・キャッシュメモリ方式は、ダイレクトマップ方式の1次キャッシュメモリと小規模なフルアソシアティブ方式の2次キャッシュメモリを用いて、高速かつ低ミス率のキャッシュシステムを構成する。

【0031】これに対し、本発明は、ダイレクトマップ方式の1次キャッシュメモリと小規模なダイレクトマップ方式の2次キャッシュメモリを用いて、高速かつ低ミス率かつ制御が容易なキャッシュシステムを構成するという独創的内容を有する。

【0032】

【実施例】次に図面を用いて、本発明の実施例を説明す

る。

【0033】（実施例1）図1は、本発明の実施例1を示すブロック図である。

【0034】図1において、100はマイクロプロセッサLSI、110はCPU（中央処理装置）、120は1次キャッシュメモリ、140は2次キャッシュメモリ、160はメインメモリ（主記憶装置）である。ここに、CPU110、1次キャッシュメモリ120、2次キャッシュメモリ140を同一のマイクロプロセッサLSI100に集積し、マイクロプロセッサ装置を構成する。

【0035】111はバイト単位のメモリアドレスとアクセスするデータの大きさを指定可能で、CPU110が出力するアドレスバス、112はCPU110が入出力するデータバス、113は1次キャッシュメモリ120が出力するバス待合わせ信号、131は1次キャッシュメモリ120が出力するアドレスバスである。

【0036】また、132は1次キャッシュメモリ120が入出力するデータバス、136はメインメモリ160のデータを1次キャッシュメモリ120に伝えるバイパス、133は2次キャッシュメモリ140が出力するバス待合わせ信号、151は2次キャッシュメモリ140が出力するアドレスバス、152は2次キャッシュメモリ140が入出力するデータバス、153はメインメモリ160が出力するバス待合わせ信号である。

【0037】121、141はアドレス111、131の下位ビット（インデクス）をアドレスとしてアドレス111、131の上位ビット（タグ）を格納する2ⁿ語×tビットのタグ用メモリ、122、142はアドレス111、131の下位ビット（インデクス）をアドレスとして各エントリが有効かどうかを示す2ⁿ語×2ビットのステータスフラグ用メモリである。

【0038】123、143はアドレス111、131の下位ビット（インデクスとオフセット）をアドレスとしてデータバス112、132またはデータバス132、152の値を読み書きする2^(1+b-2)語×32ビットのデータ用メモリ、124、144はアドレス111、131の上位ビット（タグ）とタグ用メモリ121、141の出力が一致しているか検出する比較器である。

【0039】125、145は比較器124、144、ステータスフラグ用メモリ122、142の出力とアドレス111、131からキャッシュメモリ120、140を制御する制御回路である。

【0040】126、146はキャッシュメモリ120、140がミスしたときにメモリアドレスを生成するアドレスバッファ、134はタグ用メモリ121のタグ情報を一時保持してタグ用メモリ141に伝えるタグバッファである。

【0041】135はデータ用メモリ123のデータを

7

一時保持してデータ用メモリ143に伝えるタグバッファである。

【0042】アドレスバス111, 131のアドレス値がどのようにキャッシュメモリ120, 140で使われるかを図2に示す。ブロックサイズが 2^b バイトとすると、ブロック内オフセットは b ビット幅、キャッシュのエントリ数が 2^i エントリとすると、インデクスは i ビット幅、タグはアドレスのビット幅からインデクスとオフセットを引いた幅をもつ。

【0043】例えば、32ビットアドレスで、1次キャッシュの容量が8Kバイト、ブロックサイズが16バイトの場合、ブロック内オフセットは4ビット($2^4 = 16$)、エントリ数は512($= 8K / 16$)であるので、インデクスは9ビット($2^9 = 512$)、タグは19ビット($= 32 - 9 - 4$)となる。

【0044】ブロックサイズ b は、1次キャッシュメモリ120と2次キャッシュメモリ140で同じ大きさとする。

【0045】次に、図1のキャッシュメモリ方式の動作について、図3, 図4を用いて説明する。

【0046】制御回路125によって制御される、1次キャッシュメモリ120の動作を図3に示す。ステータスフラグ122の内容は「無効」に初期化されているものとする。

【0047】CPU110によるメモリアクセスが発生すると、アドレス111のインデクスをアドレスとしてメモリ121, 122を読み出し(ステップ301)、タグ用メモリ121の出力とアドレス110のタグを比較器124で比較する(ステップ302)。比較が等しく、かつ、ステータスフラグ122が「有効」または「書込」を示す場合はヒット、それ以外はミスである。

【0048】ヒットした場合、アドレス111のインデクスおよびブロック内オフセットをアドレスとして、データバス112を介して、データ用メモリ123の内容を読み書きする(ステップ304, 305)。

【0049】CPU110からのアクセスが書込みである場合は、データ用メモリ123の内容を、エントリ中の指定されたアドレスに指定されたサイズ(バイト幅、ワード幅など)の部分だけ書換え、ステータスフラグに「書込」を書込む(ステップ305, 306)。

【0050】ミスした場合、まず、ステータスフラグ122が「書込」か「有効」か「無効」かを判定し(ステップ307)、「書込」ならばインデクスで指定されているエントリを2次キャッシュメモリ140に書込む(ステップ308)。このとき、タグメモリ122から読み出した値をアドレスとし、データはデータ用メモリ123の内容である。ステータスフラグ122が「有効」ならば該当するエントリのタグとデータをバッファ134, 135に保存する(ステップ309)。

【0051】CPU110からのアクセスが読み出しの

8

場合は、次に、外からアドレス111で指定されたエントリを読み込む。読み込んだエントリに対して、アドレス111で指定されたサイズ分だけ、データ用メモリ123の内容をデータバス112を介してCPU110に転送する(ステップ311)。同時に、タグ用メモリ121にアドレス111のタグを、ステータスフラグ122に「有効」を書込む(ステップ312, 313)。

【0052】CPU110からのアクセスが書込みの場合は、次にタグ用メモリにアドレス111のタグを書込む(ステップ314)。同時に、データ用メモリ123の内容を、エントリ中の指定されたアドレスに指定されたサイズ(バイト幅、ワード幅など)の部分だけが書換え、ステータスフラグに「書込」を書込む(ステップ305, 306)。

【0053】制御回路145によって制御される、2次キャッシュメモリ140の動作を図4に示す。動作は、1次キャッシュメモリ120と似ているが一部異なる。

【0054】1次キャッシュメモリ120によるメモリアクセスが発生すると、アドレス131のインデクスをアドレスとしてメモリ141, 142を読み出し(ステップ401)、タグ用メモリ141の出力とアドレス131のタグを比較器144で比較する(ステップ402)。比較が等しく、かつ、ステータスフラグ142が「有効」または「書込」を示す場合はヒット、それ以外はミスである。

【0055】ヒットした場合、アドレス131のインデクスおよびブロック内オフセットをアドレスとして、データバス132を介して、データ用メモリ143の内容を読み書きする(ステップ404, 408)。1次キャッシュメモリ120からのアクセスが読み出しである場合は、バッファ134, 135に保存してある1次キャッシュメモリ120の該当エントリの内容をメモリ141, 143に転送する(ステップ405, 406)。

【0056】それから、ステータスフラグ142に、読出しアクセスなら1次キャッシュメモリ120の該当エントリのステータス値を、書込アクセスなら「書込」を書込む(ステップ407, 409)。

【0057】ミスした場合、まず、ステータスフラグ142が「書込」か「有効」か「無効」かを判定し(ステップ407)、「書込」ならばインデクスで指定されているエントリをメインメモリ160に書込む(ステップ411)。

【0058】このとき、タグメモリ142から読み出した値をアドレスとし、データはデータ用メモリ143の内容である。外へのアクセスは1エントリ分であり、ブロックサイズがバス幅より大きいときは複数のバスサイクルとなる。

【0059】1次キャッシュメモリ120からのアクセスが読み出しの場合は、次に、メインメモリからアドレス131で指定されたエントリを読み込み、2次キャッ

シユメモリの中を素通り(152, 136, 132)させて1次キャッシュメモリ120にデータを転送する。

【0060】1次キャッシュメモリ120からのアクセスが書込みの場合は、次に、タグ用メモリ141にアドレス131のタグを書込む(ステップ414)。同時に、データ用メモリ143の内容を、エントリ中の指定されたアドレスに指定されたサイズ(バイト幅、ワード幅など)の部分だけが書換え、ステータスフラグに「書込」を書込む(ステップ408, 409)。

【0061】1次キャッシュメモリ120がミスした場合の処理が終了するまで、制御回路125は信号113によってCPU110を待たせておく。同様に、2次キャッシュメモリ140へのアクセス(ステップ408, 410)は、制御信号133によって待たされる場合がある。

【0062】キャッシュメモリ単体の制御方式自体は公知のものである。

【0063】ここで述べたキャッシュメモリ単体の制御方式は、一般にはライトバック(コピーバック、ストアインとも呼ぶ)という方式である。ステータスフラグ122, 142は、3状態(「無効」, 「有効」, 「書込」)を持つ。上位からの書込みアクセスは、直接には下位に伝えられない。

【0064】上位からの書込みアクセスがあると、直接、下位のメモリを書込む方式(ライトスルーと呼ぶ)もある。ステータスフラグ122, 142は、2状態(「無効」, 「有効」)しかなく、制御は単純である。しかし、ライトバック方式よりミス率は高い。

【0065】マイクロプロセッサに対応するために、ステータスフラグの状態を更に増やしたライトバック方式もある。

【0066】通常のライトバック方式と異なるのは、1次キャッシュメモリと2次キャッシュメモリの間にデータの重複がないようにしていることである。

【0067】読み出しアクセスにおいて、1次、2次キャッシュメモリ共にミスのときは、メインメモリから1次キャッシュメモリにデータが転送され、1次キャッシュメモリがミスし2次キャッシュメモリがヒットしたときは、1次キャッシュメモリと2次キャッシュメモリのデータを入替える。このように制御することで、1次キャッシュメモリと2次キャッシュメモリに同一アドレスのデータは重複しない。よって、2次キャッシュメモリの効率を上げている。

【0068】本発明の要旨は、キャッシュメモリ単体の制御方式でなく、階層構造と、1次/2次キャッシュメモリ *
相対性能 = 1 / 平均アクセス時間

ここで、

M1 1次キャッシュメモリのミス率

M2 2次キャッシュメモリのミス率

B1 1次キャッシュメモリのアクセス時間 (clock 50

*メモリを同等のアクセス時間でアクセスできること、1次と2次のキャッシュメモリ間での転送が高速にできることである。

【0069】これらのキャッシュメモリ単体の制御方式が、キャッシュメモリ120, 140にどのような組合せで実現されていても、本発明の主旨に見合うことは言うまでもない。

【0070】(実施例2)次に本発明の実施例2を説明する。図5に本発明の実施例2を示す。実施例1(図1)と異なるのは、2カ所ある。ひとつは、2次キャッシュメモリ140において、メモリ141, 142, 143のアドレスとなるインデクスを、インデクス生成回路146を用いて生成していることである。もうひとつは、タグ用メモリの幅(ビット)が、実施例1より広いことである。

【0071】アドレスバス131のアドレス値がどのようにキャッシュメモリ140で使われるかを図6, 図7に示す。図2と違って、アドレスからブロック内オフセットを抜いた分が、タグとなっている。インデクスは、タグの一部をインデクス生成回路146によってハッシュして得る。ハッシュ関数としては、図6, 図7のように、階層的に排他的論理和をとったものを用いる。

【0072】アドレスのハッシュされる部分は、1次キャッシュメモリのインデクスとタグの一部(アドレスのビット位置で)を含むようにすると良い。コンフリクトミスが解消されミス率が減るためである。

【0073】図6の例ではハッシュするビットを下位から詰めて取っているが、図7のようにタグ全体から取ってきてもよい。

【0074】また、図6, 図7では排他的論理和を2段階のツリー構造にしているが、1段階だけにしても、3段階以上のツリーにしてもよいことは言うまでもない。

【0075】インデクス生成回路146によって2次キャッシュメモリ140のアクセス時間は長くなる。しかし、CPU110から1次キャッシュメモリ120へのアクセスと並行して、インデクス生成回路146を動作させることによって、アクセス時間を長くしないことができる。

【0076】ここで、従来例、実施例の性能とハードウェアコストを計算してみる。

【0077】ノーウェイトのメモリを毎クロックアクセスできるときの性能を1.0とした、CPUの相対性能は以下のように表わせる。CPUからは毎クロックでメモリアクセスが発生するものとする。

【0078】

... (1)

k)

B2 2次キャッシュメモリのアクセス時間 (clock k)

B3 メインメモリのアクセス時間 (clock)

11

12

b ブロックサイズ (Byte)

* 【0079】 キャッシュメモリが1階層 (従来方式) の場合、

a バス幅 (Byte)

とすると、

*

$$\text{相対性能} = 1 / \{ (1 - M1) \times B1 + (M1 \times b / a \times B3) \} \quad \dots (2)$$

【0080】 キャッシュメモリが2階層 (ヴィクティム キャッシュまたは本発明) の場合、

$$\text{相対性能} = 1 / \{ (1 - M1) + (M1 \times (1 - M2) \times b / a \times B2) + (M1 \times M2 \times b / a \times B3) \} \quad \dots (3)$$

【0081】 ここで、例を挙げる。

10

$$B1 = 1 \text{ clock}, B2 = 1 \text{ clock}, B3 = 8 \text{ clock} \\ b = 16 \text{ Byte}, a = 4 \text{ Byte} \quad \dots (4)$$

とする。

※示す。相対性能は式 (2), (4) から求めた。

【0082】 あるアプリケーション (diff: 16KB 程度のデータ) において、ダイレクトマップ方式のキャッシュメモリの容量とミス率を測定した結果を表1に※

【0083】

【表1】

	容 量	ミ ス 率	相 対 性 能
例 1	8 Kバイト	2. 12%	0. 570
例 2	16 Kバイト	1. 59%	0. 644

【0084】 同じアプリケーションで、ヴィクティムキャッシュ方式、実施例1, 2について測定した結果を表2に示す。相対性能は、式 (3), (4) から求めた。1次キャッシュメモリは全て、8 Kバイトのダイレクトマップ方式 (表1の例1と同じ) である。

★の実施例2の例である。

【0086】 上記6例の性能を、図8にまとめる。さらに、図8には、各例をLSI化したときの相対面積 (例2を1とした) と、効率 (相対性能/相対面積) を示した。

【0085】 例3はヴィクティムキャッシュの一例、例30 4, 例5は本発明の実施例1の例、例6, 例7は本発明★

【0087】

【表2】

	2次キャッシュ方式	容 量	ミス率	相 対 性 能
例3	フルアソシアティブ	4エントリ	60. 7%	0. 634
例4	ダイレクトマップ	8エントリ	91. 0%	0. 560
例5	ダイレクトマップ	128エントリ	85. 9%	0. 574
例6	ハッシュ付ダイレクトマップ	8エントリ	60. 0%	0. 635
例7	ハッシュ付ダイレクトマップ	128エントリ	27. 3%	0. 757

【0088】

【発明の効果】 以上説明したように、本発明によるキャッシュメモリ方式は、以下の効果を有する。

①性能が高い。図8から分かるように、本発明の性能は 50

高い。実施例 (例7) 2では、例2 (1次キャッシュメモリの容量を倍にしたもの) よりも性能が高い。

②ハードウェア効率が良い。図8から分かるように、本発明によるハードウェア量 (LSI化したときの面積)

13

は、いずれも例 2 (1 次キャッシュメモリの容量を倍にしたもの) よりも小さい。例 6 (実施例 2) では、ヴィクティムキャッシュ (例 3) より小面積で、同等の効率を実現している。例 3 ではフルアソシアティブ方式を用いるため、面積効率が非常に悪い。例 7 では、8 K バイトの 1 次キャッシュメモリに 2 K バイトの 2 次キャッシュメモリを付加しただけで、例 2 (16 K バイトの 1 次

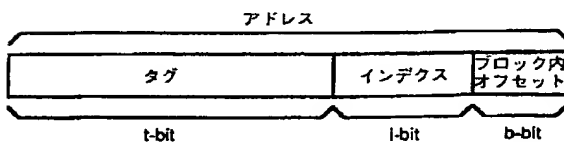
キャッシュメモリ) よりも高い性能を実現している。
③ハードウェアが簡単である。従来例で示したように、ヴィクティムキャッシュ方式の 2 次キャッシュメモリで用いられるフルアソシアティブ方式には、制御論理が複雑であるという欠点がある。ミスした場合にどのエントリを入替えるかを定める制御論理 (通常は LRU: Least Recent Used) は複雑であり、通常は比較器を介してデータメモリをアクセスするが、入替え時にはランダムアクセスできなければならない。それゆえ、テストも難しい。

【0089】ところが、本発明による 2 次キャッシュメモリはダイレクトマップ方式であり、非常にシンプルである。ハードウェアは SRAM+比較器ひとつという単純さである。また、1 次キャッシュメモリと 2 次キャッシュメモリが同じ方式であるため、制御回路の設計が容易 (使い回しができる) である。

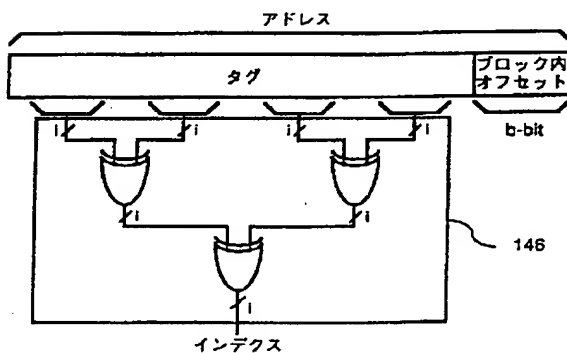
【図面の簡単な説明】

【図 1】本発明の実施例 1 のハードウェア構成を示す図

【図 2】



【図 7】



14

である。

【図 2】ダイレクトマップ方式のアドレスの構成を示す図である。

【図 3】本発明の動作を説明する図である。

【図 4】本発明の動作を説明する図である。

【図 5】本発明の実施例 2 のハードウェア構成を示す図である。

【図 6】実施例 2 のアドレスの構成を示す図である。

【図 7】実施例 2 のアドレスの構成を示す図である。

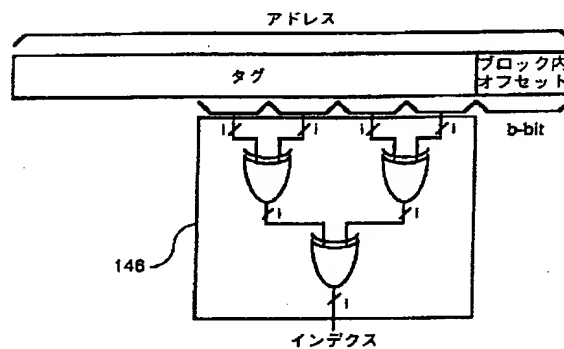
【図 8】従来のキャッシュメモリ方式のミス率を示す図である。

【図 9】本発明の効果を示す図である。

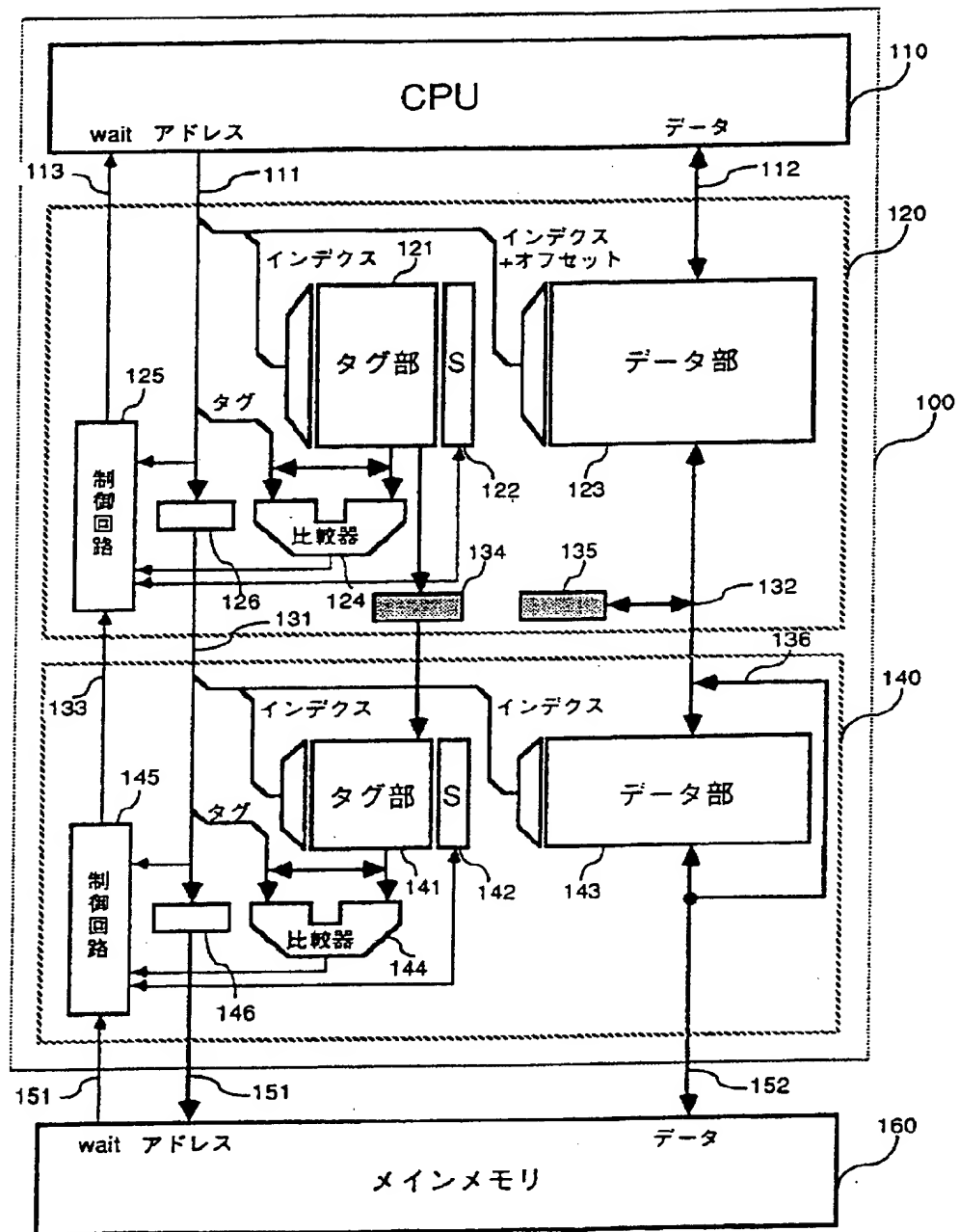
【符号の説明】

100 マイクロプロセッサ LSI
110 CPU
120 1 次キャッシュメモリ
140 2 次キャッシュメモリ
160 メインメモリ
121, 141 タグ用メモリ
122, 142 ステータスフラグ用メモリ
123, 143 データ用メモリ
124, 144 比較器
125, 145 制御回路
146 インデクス生成回路

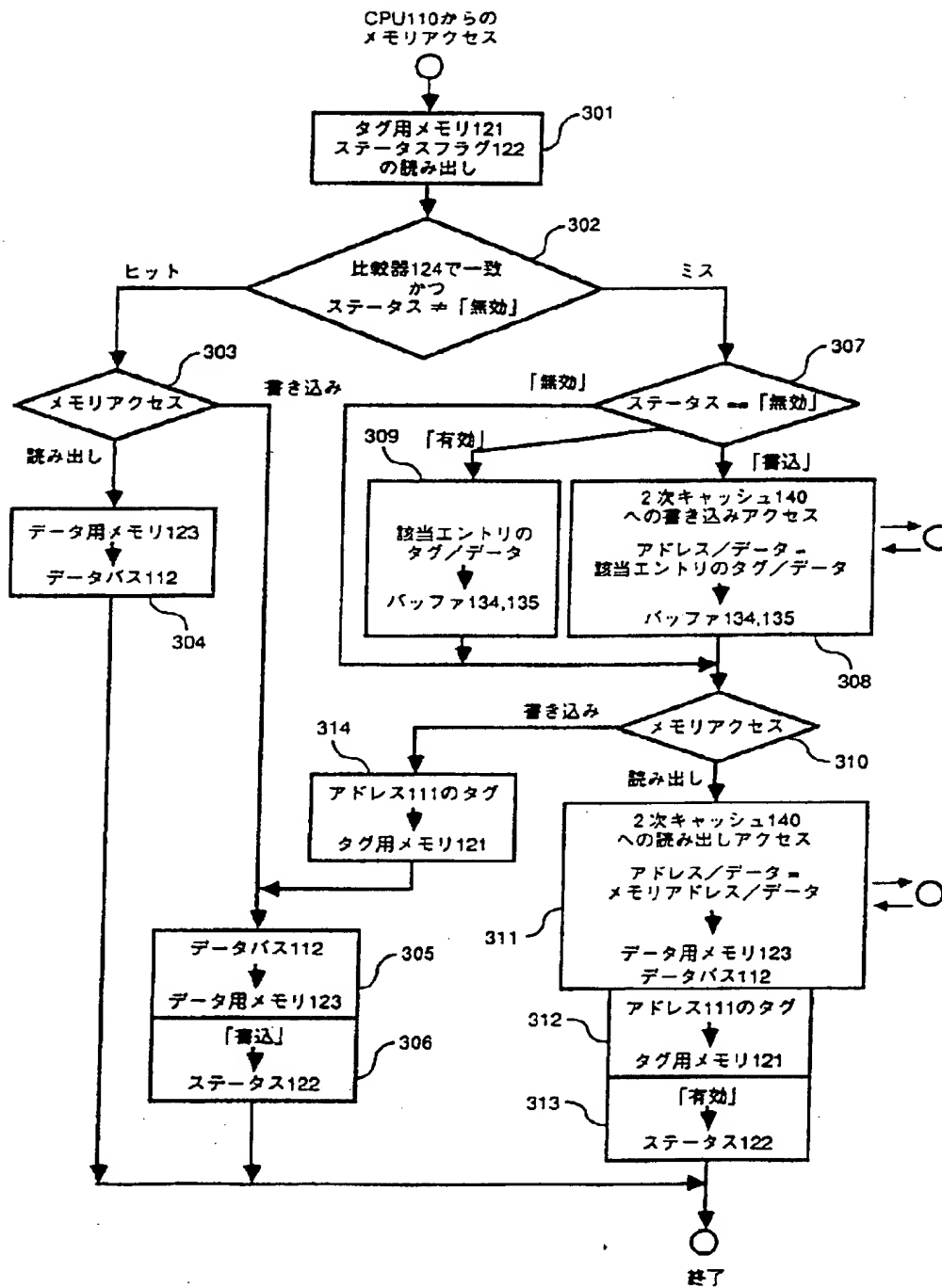
【図 6】



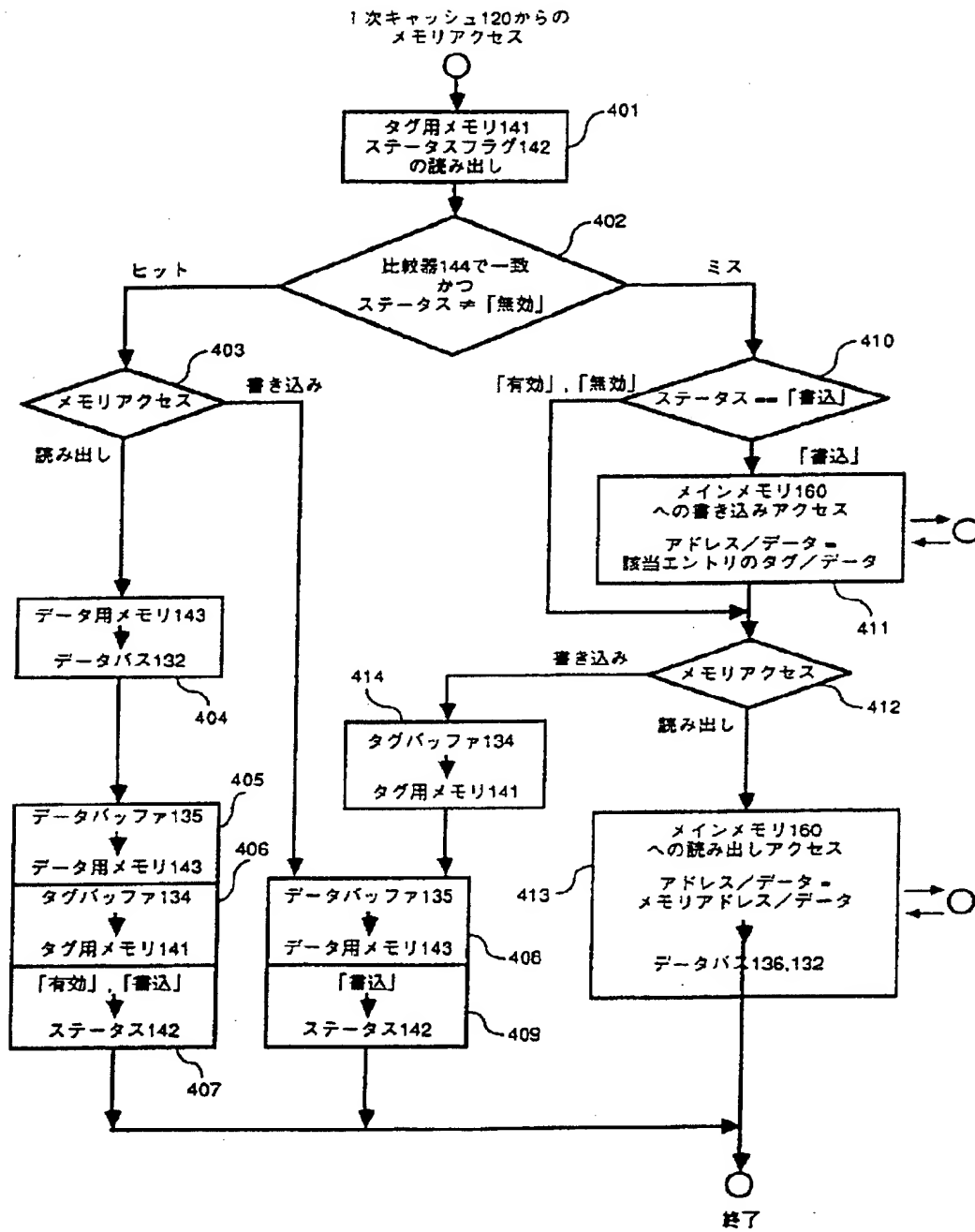
【図 1】



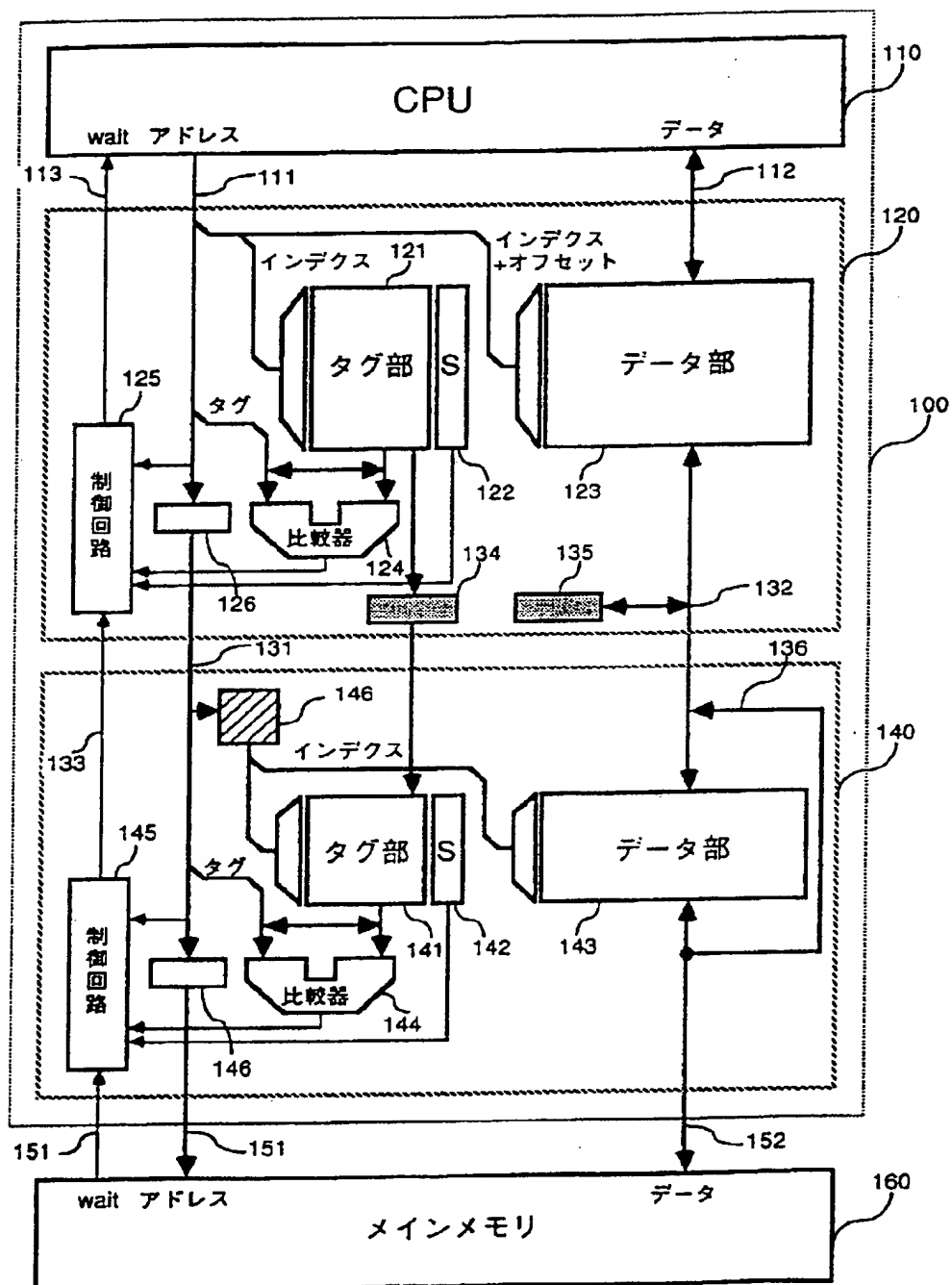
【図3】



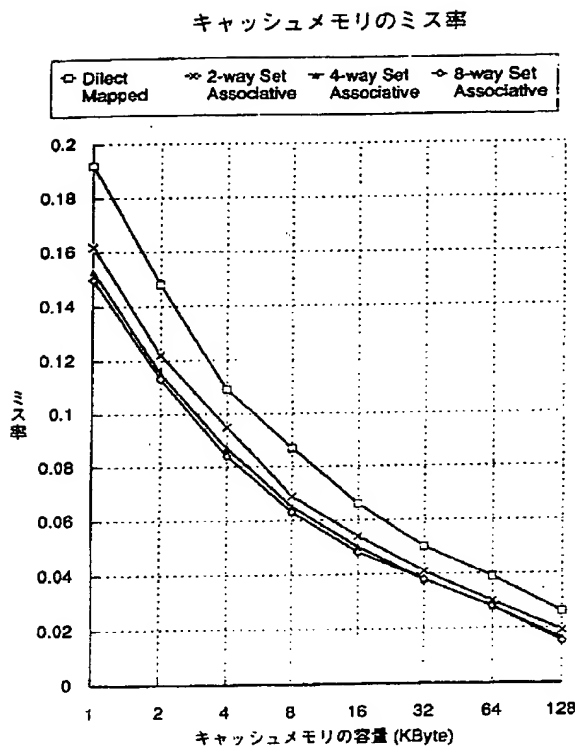
【図4】



【図5】



【図8】



【図9】

